

Climate and Forecasting (CF) Conventions

7th Joint Earth Science Data System Working Group
Philadelphia

V. Balaji¹ John Caron² Jonathan Gregory³ Steve Hankin⁴
Bryan Lawrence⁵ Russ Rew² Rich Signell⁶ Karl Taylor⁷

¹Princeton University

²Unidata

³University of Reading

⁴NOAA/PMEL

⁵British Atmospheric Data Centre

⁶United States Geological Survey

⁷Program for Climate Model Diagnosis and Intercomparison

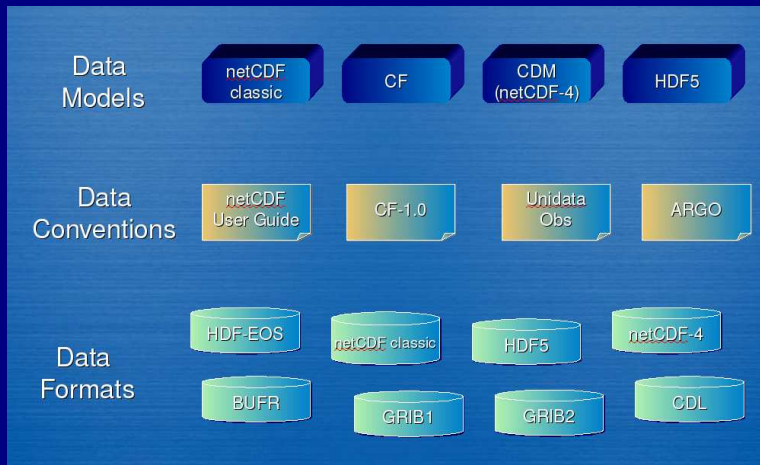
Talk outline...

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

Formats, Conventions, Data Models



What are the CF Conventions?

- A standard for encoding Climate and weather Forecast metadata in netCDF files: `cfconventions.org`
- Metadata conventions supporting interoperability for earth science data from different sources
- Intended for both model output and observational datasets
- Examples of CF metadata
 - Coordinate information needed to locate data in space and time
 - Standard names for quantities to determine whether data from different sources are comparable
 - Additional grid information (e.g., grid cell bounds, cell averaging methods)

Goals of the CF Conventions

- Locate data in space-time and as a function of other independent variables, to facilitate processing and graphics
- Identify data sufficiently to enable users of data from different sources to decide what is comparable, and to distinguish variables in archives
- Framed as a netCDF standard, but most CF ideas relate to metadata design in general and not specifically to netCDF, and hence can be contained in other formats such as XML
- Backwards-compatible with prior standards: COARDS is a fully-contained subset, though some COARDS features are deprecated.

CF data descriptors

- Data provenance: title, institution, contact, source (e.g model), history (audit trail of operations), references, comment
- Description of associated activity: project, experiment
- Description of data: units, standard_name, long_name, auxiliary_variables, missing_value, valid_range, flag_values, flag_meanings
- Description of coordinates: coordinates, bounds, grid_mapping (with formula_terms); time specified with reference_time (“time since T0”) and calendar attributes.
- cell_methods, cell_measures, and climatological statistics.

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

Where is CF Metadata used?

- Widely used and accepted in the climate community
 - World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset, used by Intergovernmental Panel on Climate Change (IPCC) Working Group 1
 - Paleoclimate Modeling Intercomparison Project (PMIP), Hemispheric Transport of Air Pollution (HTAP), regional groups, EU-funded ENSEMBLES prediction system for climate change, ...
 - Planned use in model archives for next IPCC cycle (CMIP5/AR5)
- Widely adopted in other netCDF archives for atmosphere, oceans, and surface data: ESMF, GFDL, Hadley Centre, NCAR, NOAA, ...
- Supported by various software packages with facilities for analyzing, visualizing, subsetting, regridding, and aggregating data

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

A brief history of CF

- Evolved from simple netCDF User Guide conventions (1989), COARDS standard (1995), GDT (1999), and NCAR CSM (1999) conventions
- 2000-2003: Developed by volunteer efforts (Brian Eaton, Jonathan Gregory, Bob Drach, Karl Taylor, and Steve Hankin)
- 2003: CF 1.0 released
- 2005: CF white paper discussing future governance circulated
- 2006: Revised white paper presented to WCRP WGCM
- 2007: Rules for community-initiated changes to CF conventions agreed upon

Guiding Principles of CF

- Data should be self-describing, without external tables needed for interpretation.
- Conventions should only be developed for things we know will be needed.
- Conventions should not be onerous to use for either data-writers or data-readers.
- Metadata should be readable by humans as well as easily interpretable by programs.
- Redundancy should be minimized to avoid inconsistencies when writing data.

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

CF Governance Structure

- CF Governance Panel established
- Control turned over to two working committees:
 - CF Conventions
 - CF Standard Names
- Committee work done via email and archived web discussion at `cfconventions.org`
- WCRP/WGCM has been asked to assume responsibility for stewardship
- WCRP/WGNE has been invited to appoint representation on CF Governance Panel

Strengths of CF Governance

- Successful international collaboration to codify best practices into a community standard
- Proven record of achieving interoperability
- Engagement of diverse communities to capture expertise for standard names
- Agreement on open process for evolving conventions and reaching consensus
- Commitment of organizational infrastructure and resources
 - BADC: Standard names – (50% FTE)
 - LLNL PCMDI: Web site support – (20% FTE)
 - UCAR Unidata: Library development (`libcf`) – (10% FTE)
- Discussion of CF issues at annual GO-ESSP (Global Organization for Earth System Science Portals) meetings:
<http://go-essp.gfdl.noaa.gov>

Issues with CF Governance

- How to get volunteers from community to help with
- Creating and reviewing proposals to address new technical issues
- Testing adequacy of proposed extensions
- How to balance desired simplicity versus necessary complexity?
- How to balance immediate needs of data providers versus stability needed by application developers?
- How to resist temptation to tinker, oversimplify, or over-generalize?
- Peaceful co-existence with other standards: OGC, MMI, . . .

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

Future directions of CF

- Implementing CF metadata conventions for other file formats (besides netCDF)
- Supplying both data providers and application developers with library support for using CF
- Providing improvements for representing observational data and metadata
- Supporting more types of grids (staggered, curvilinear, nested)
- Supporting mappings between CF and other metadata standards and conventions
- Use of netCDF-4 data model and format

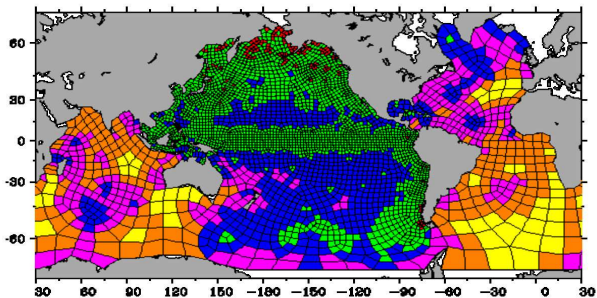
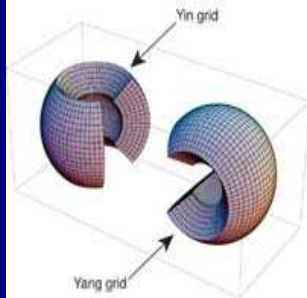
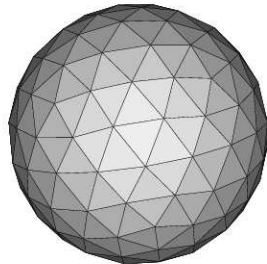
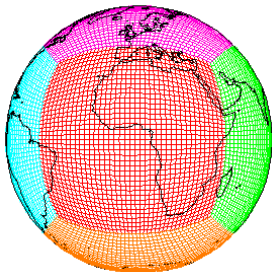
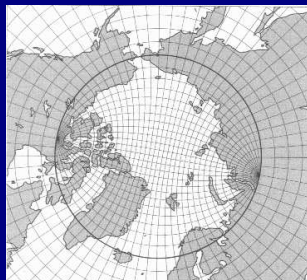
Some current hot ticket items in CF

Several complex proposals and projects are currently under active discussion on the lists:

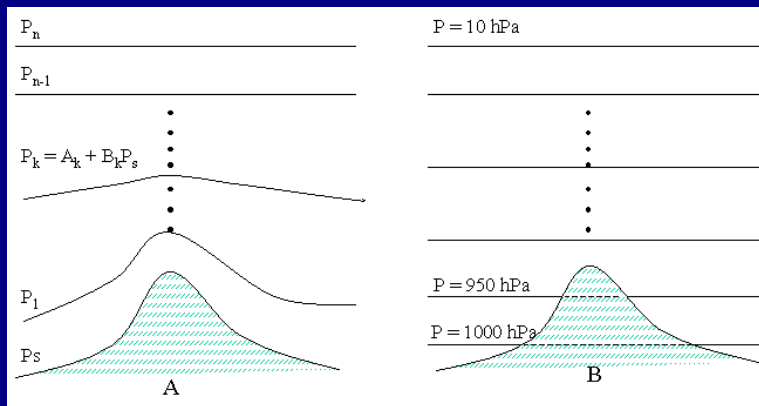
- *Ensemble axis*: Representation of an ensemble of model runs.
- *Aggregation*: XML representations of netCDF (e.g NCML) allow an aggregated “dataset” view of sets of files.
- *Semantic mediation*: adding more semantic-web ideas to CF to reconcile differing vocabularies, express synonymy, supersession, subsumption.
- *Grid specification*: A richer syntax for grids to express mosaics, unstructured grids, etc.

<http://www.gfdl.noaa.gov/~vb/gridstd/gridstd.html>

Horizontal grids in use in ESMs



Vertical coordinates



The taxonomy of vertical coordinates distinguishes **mass-based** and **space-based** vertical coordinates. There is often an attempt to do something in the spirit of geo-referencing: invoking a “standard” reference grid: usually based on pressure levels in the atmosphere, and depth in the ocean.

Talk outline . . .

- 1 What are the CF Conventions?
- 2 How widely used is CF metadata?
- 3 How were the CF Conventions developed?
- 4 How is CF governed?
- 5 Where is CF headed?
- 6 What's on the "Concluding Comments" slide?

Concluding comments

- CF has undergone a two-year transition from informal maintenance by its authors to community governance.
- The CF Conventions transition seems moderately successful so far, but needs more active engagement by community volunteers.
- The CF Standard Names transition is also successful, with over 50 contributors and 900 standard names.
- Wide usage and real-world experience suggests CF metadata conventions are highly suitable for a broad community of data providers and users.
- To guarantee maintenance and ensure persistence as an internet resource, CF will need either
 - a single recognized authoritative organization to provide stewardship, or
 - a continued supply of interested and knowledgeable volunteers